FB9402V2

ECS Science Requirements Summary

White Paper

Working Paper

March 1994

Prepared Under Contract NAS5-60000

RESPONSIBLE ENGINEER

Andrew Endal, Deputy Project Scientist Date
EOSDIS Core System Project

SUBMITTED BY

Robert Curran, Science Office Manager Date
EOSDIS Core System Project

Hughes Applied Information Systems, Inc.
Landover, Maryland

This page intentionally left blank.

Contents

1. Introduction

2. ECS Science Requirements

Figures

Tables

Abbreviations and Acronyms

1. Introduction

1.1 Purpose

This paper provides a science-oriented summary of the ECS requirements to provide a vehicle for communication between the ECS project and the science community. Since the readership will include scientists with only a casual familiarity with the ECS project, the paper provides significant background material and avoids project-specific terminology where possible.

Dr. Steven Wharton (NASA EOSDIS Project Scientist) contributed material for a previous incarnation of this paper. He and many staff members of the ECS Project provided useful comments on draft versions.

1.2 Organization

Section 2.1 summarizes the EOS science objectives, while Section 2.2 describes the end-to-end flow of science data from the EOS instruments to the science user. These sections are intended as background material. The ECS requirements are summarized in Section 2.3, from a science user perspective. A high-level view of the ECS development process is provided in Section 2.4. Section 2.5 provides an annotated list of addition source materials on the ECS requirements, design, and development process. Complete references for cited papers are given in Section 2.6.

1.3 Review and Approval

This White Paper is an informal document approved at the Office Manager level. It does not require formal Government review or approval; however, it is submitted with the intent that review and comments will be forthcoming.

Questions regarding technical information contained within this Paper should be addressed to the following ECS and/or GSFC contacts:

- ECS contact

Andrew Endal, ECS Deputy Project Scientist
andy@eos.hitc.com
(301) 925-0750

- GSFC contact

Steven Wharton, EOSDIS Project Scientist
wharton@eosdata.gsfc.nasa.gov
(301) 286-9394

Questions concerning distribution or control of this document should be addressed to:

Data Management Office
The ECS Project Office
Hughes Applied Information Systems, Inc.
1616A McCormick Dr.
Landover, MD 20785

2. ECS Science Requirements

2.1 ECS Objectives

Mission to Planet Earth (MTPE) is a long-term, multi- and inter-disciplinary NASA research mission to study the processes leading to global change and develop the capability to predict the future evolution of the Earth system on time scales of decades to centuries. To accomplish these objectives, researchers require a readily accessible collection of diverse observations of the Earth over an extended period of time, with the capability to create and add new data products to this collection based on improved understanding. MTPE aims not just at a study of the disciplinary sciences of the atmosphere, oceans, cryosphere, biosphere, and solid Earth, but at the inter-disciplinary interactions among these often disparate realms of study. This is necessary for development of a predictive capability for modeling the Earth system as the scientific basis for global environmental policy.

MTPE comprises three major components:

- the Earth Observing System (EOS) will collect Earth science data, with emphasis on long-term, sustained data sets from carefully calibrated instruments on satellites in low Earth orbit,
- the EOS Data and Information System (EOSDIS) will provide the Earth science community with easy, affordable, and reliable access to EOS and other Earth science data, and
- an integrated scientific research program will investigate processes in the Earth system and use this information to improve predictive models.

As a byproduct, the EOSDIS will provide a broader community of users with a unique resource for enhancing their understanding of global change issues and for acquiring data for use in other applications.

The EOSDIS Core System (ECS) is the major component of the EOSDIS. The ECS will control the EOS spacecraft and instruments, process data from the EOS instruments, and manage and distribute EOS data products and other selected data sets (see Table 2-1). Interoperating with other data systems maintained by government agencies and the research community, the ECS will provide comprehensive services for accessing Earth science data.

A goal of the ECS program is to provide a highly adaptable system that is responsive to the evolving needs of the Earth science community. Over the system lifetime (at least two decades beyond the launch of the first EOS spacecraft), evolution will come from at least three separate sources:

- scientific needs will change as Earth system science matures and new applications of the data emerge,

*Table 2-1. ECS Data Sources**

Mission	Launch	Objectives	ECS role† f	p	a	d
		Pathfinder data sets §				
NOAA-AVHRR	ongoing	land surface radiances & vegetation, sea surface temperature, global cloud & aerosol properties			•	•
GOES	ongoing	cloud imagery and radiation budget				•
Landsat 4/5	ongoing	land surface / vegetation mapping				•
DMSP-SSM/I	ongoing	precipitation, wind speed & liquid water over oceans, sea ice, surface temperature and vegetation density			•	•
NOAA-TOVS	ongoing	atmospheric temperature / humidity profiles, cloud fractions & surface pressure			•	•
		International Partner platforms				
ERS-1	1992	high latitude ice & ocean surface conditions				•
JERS-1	1992	high latitude ice & ocean surface conditions				•
TOPEX/ Poseidon	1993	ocean surface & ice sheet topography			•	•
ERS-2	1994	high latitude ice & ocean surface conditions				•
ADEOS	1995	vis/IR imagery & ocean surface conditions			•	•
Radarsat	1995	high latitude ice & ocean surface conditions				•
ADEOS II	1999	vis/IR imagery & ocean surface conditions			•	•
		Earth Probe missions				
UARS	1991	upper atmosphere chemistry and dynamics			•	•
TOMS	1993/95	ozone mapping & monitoring			•	•
SeaWiFS	1994	ocean color & biological productivity			•	•
TRMM	1997	tropical precipitation measurement			•	•
		radiation budget and lightning (CERES & LIS)		•	•	•
Landsat 7	1997	high resolution vis/IR land imagery				•
		EOS spacecraft (through 2002)				
EOS-AM	1998	physical climate, chemistry & cloud interactions	•	•	•	•
COLOR	1998	ocean color & biological productivity		•	•	•
AERO	2000	global aerosols & stratospheric chemistry	•	•	•	•
EOS-PM	2000	physical climate, hydrology & cloud interactions	•	•	•	•
ALT	2002	ocean surface & ice topography	•	•	•	•
CHEM	2002	atmospheric chemistry & sea ice altimetry	•	•	•	•

Notes: * For additional information on these missions and instruments, see the EOS Reference Handbook (Asrar and Dokken, 1993).
† f = flight operations; p = processing; a = archival; d = data access
§ Derived from reprocessing operational sensor data to produce research quality data sets. Schwaller and Andrews (1993) provide descriptions of the planned Pathfinder data sets.

- information system technologies must be refreshed as maintaining older technologies becomes more difficult and new technologies emerge in their place, and
- changes in the information infrastructure (*e.g.*, high bandwidth networking) will lead to migration of functions to take full advantage of these capabilities.

Thus, the ECS must support the "vision" of an evolving and comprehensive information system to promote effective utilization of data for research in support of the MTPE goals. The ECS must be able to accommodate growth in all of its functions, as well as the addition of new functions. The ECS must be expandable with respect to storage and processing capacity for instrument data products and algorithms. The ECS will be extended by unique capabilities developed at the individual data centers, tailored to the needs of their user communities. The ECS must be designed to promote cost-effective development of such extensions. The ECS will interoperate with data systems of other agencies associated with the U.S. Global Change Research Program, to form the Global Change Data and Information System (GCDIS), and with International Partner data systems. Interoperating with investigator data systems, the ECS will provide input data for their higher level data processing, receive products and their associated metadata for distribution and long-term management, and provide users with access to data and services provided at the investigators' facilities.

The ECS is being procured by NASA/Goddard Space Flight Center under a system development contract to Hughes Applied Information Systems, Inc. This contract describes the ECS in terms of "requirements", which are testable statements of the capabilities of the delivered data system. These baseline requirements were derived from earlier NASA studies (Butler *et al.*, 1984; Chase *et al.*, 1986), and refined through contractor studies with the participation of the EOS Investigators Working Group and its EOSDIS Science Advisory Panel. In addition to specifying baseline requirements, the ECS development contract specifies that the design must permit expansion in both capacity and functions without major redesign, as well enabling extension of the core system to accommodate other data providers, data types, and users.

Here, we provide a synopsis of the baseline requirements and the system development process to assist Earth scientists and other potential users in guiding development of the ECS, as well as in planning their scientific utilization of the data system capabilities. ECS development must be accomplished in cooperation with the user community, with a shared commitment to the vision of an information system which promotes effective utilization of data across the entire Earth science community. This synopsis will periodically updated as the requirements are refined to capture this vision.

The ECS is part of a much larger data collection and research enterprise. In Section 2.2, the major components of this enterprise are defined to set the context for the ensuing discussion. The baseline ECS requirements are set forth in Section 2.3, in terms of services to be provided to scientific users. The ECS development process must provide flexibility to accommodate changes in user needs and incorporate new data system technologies, while satisfying cost and schedule constraints. Section 2.4 summarizes the evolutionary development process, including procedures for scientific review and feedback.

2.2 ECS Context

The observing system includes NASA instruments on satellites to be launched by NASA, the European Space Agency (ESA), and the Japanese National Space Agency (NASDA). Figure 2-1 provides a schematic view of the flow of science data from these various platforms to the users. The ECS consists of the shaded portions of Figure 2-1, plus facilities for operation of the NASA EOS satellites and instruments, including NASA instruments on International Partner satellites.

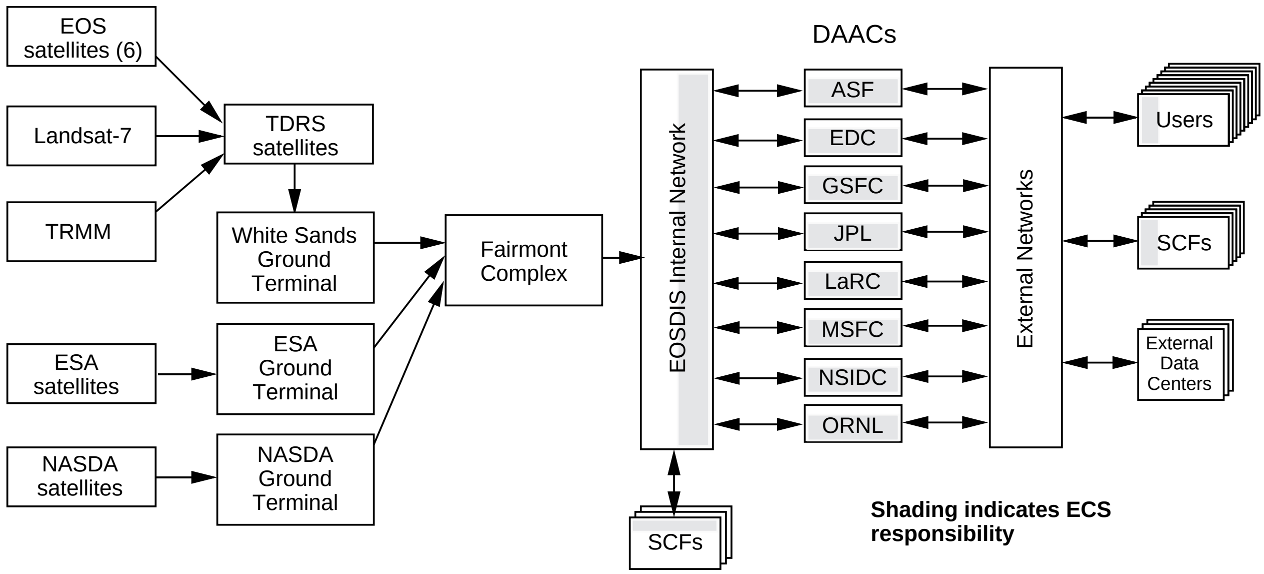

Figure 2-1. Schematic Representation of Satellite-to-user Data Flow

NASA satellites will transmit their data through the Tracking and Data Relay Satellites (TDRS), which will forward the data to the receiving station at White Sands, New Mexico. From White Sands, the data will be transmitted via dedicated circuits to the new Fairmont Complex (in West Virginia), where the data will be processed to recover the raw instrument data. International Partner satellites downlink directly to ground receiving stations. Data from NASA instruments on the International Partner platforms will be transmitted to Fairmont via commercial networks.

The data from each instrument will be sent from Fairmont to a Distributed Active Archive Center (DAAC). These data centers will house the ECS computing facilities and operational staff needed to produce EOS Standard Products and to manage and store EOSDIS data (see Table 2-1), as well as the associated metadata and browse data required for effective use of the data holdings.[1] The DAACs will exchange data via dedicated EOSDIS networks to support processing at one DAAC which requires data from another DAAC.

Eight DAACs have been selected by NASA based on their expertise in specific disciplines (indicated in Figure 2-2) and demonstrated long-term commitments to the corresponding user communities. These DAACs will provide the facilities and the management and operations support for the production, archive, and distribution of EOS Standard Products. At the DAACs, users can expect a level of service which would be difficult to maintain in a single data center attempting to serve the extraordinarily wide range of disciplines encompassed by the EOS

1 EOS Standard Products are operationally produced at the DAACs, on a prescribed schedule and using peer-reviewed algorithms supplied by the EOS investigators. Special Products are produced at the investigator facilities, and may be more experimental in nature or intended for more specialized applications.

program. It is important that a user interacting with any given DAAC will be able to access data from all the DAACs.

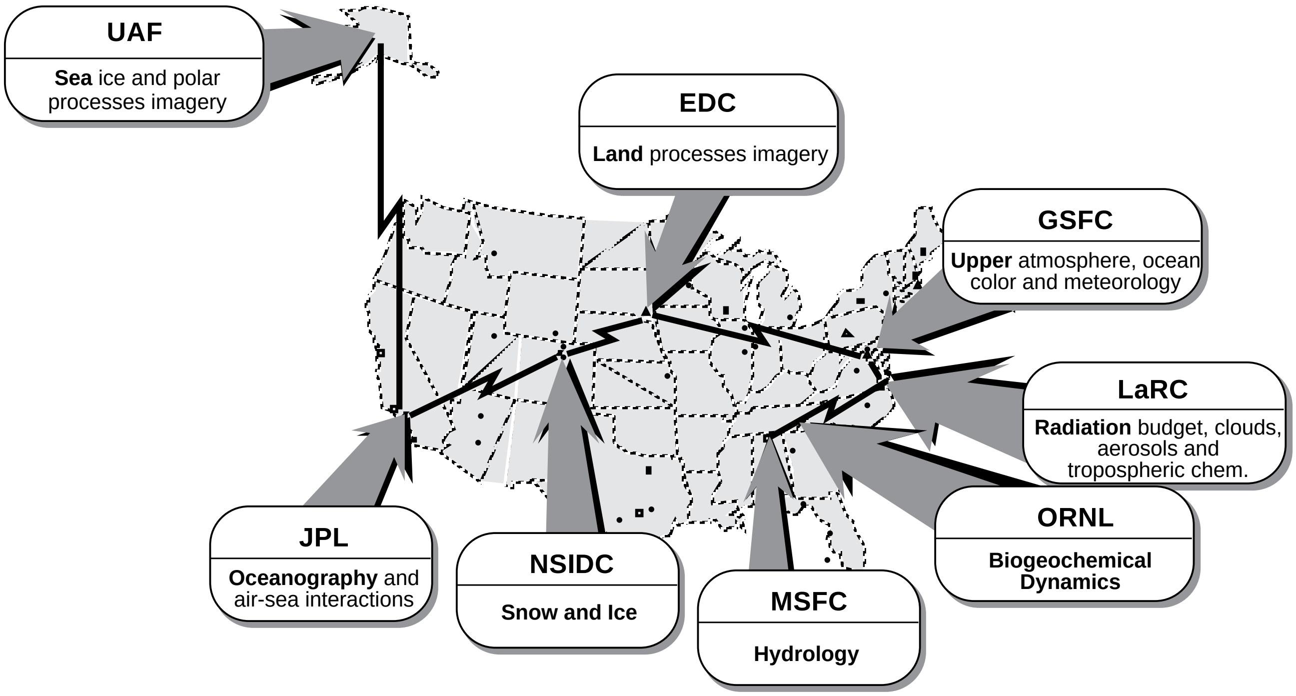

Figure 2-2. Locations and Scientific Expertise of the EOSDIS DAACs

The DAACs also house systems for processing and/or storage of non-EOS Earth science data. For example, the Alaskan SAR Facility at the University of Alaska-Fairbanks currently provides systems for receiving, processing, and archiving data from Synthetic Aperture Radars on International Partner platforms. These non-ECS activities provide valuable experience in working with both the Earth science community and the larger applications community.

Over the past several years, the DAACs have played a pivotal role in the development of the EOSDIS Version 0, integrating and evolving existing discipline data systems toward a working, integrated Earth science data system. Pathfinder data sets have been established as new, research-quality products derived from existing sensor data. Experience gained from this initial implementation of the EOSDIS distributed, inter-disciplinary system concept provides a basis for the long-term EOSDIS development and illustrates the key role to be played by the DAACs in understanding the service needs of their user communities. It is expected that DAAC-unique extensions and tailoring of standard ECS services will be provided to enhance discipline-specific data access.

Most science users will access EOS data products via shared networks (labeled External Networks in Figure 2-1). Open access to the data by all members of the science community distinguishes the EOS from previous research satellite projects, where selected investigators have had proprietary data rights for a number of years after data acquisition. It is expected that this open data policy will lead to greater utilization of EOS data products, for global change research and other applications. The EOS program is also distinguished by the large number of funded investigators (over 500), required to provide adequate expertise in the broad range of scientific disciplines in Earth system science.

Scientific Computing Facilities (SCFs), located at EOS investigators' home institutions, are used to develop and maintain algorithms (for both Standard and Special Products), calibrate the EOS instruments, validate data and algorithms, generate Special Products, provide data and services to other investigators, and analyze EOS and other data in pursuit of the MTPE science objectives. The SCFs may range from single workstations to large supercomputer data centers. While the SCFs will be developed and acquired directly by the EOS investigators, the ECS will provide software toolkits to the SCFs and other users to facilitate data access, transformation and visualization, and for algorithm development. Some SCFs will play an operational role in quality control of the EOS Standard Products; these SCFs will be linked to the DAACs via guaranteed service quality communications lines to support reliable exchange of large volumes of data.

Comprehensive understanding of Earth system processes requires data from a diverse range of sensors. Field campaign and other *in situ* data will be contributed by NOAA and by the scientific users. Remote-sensing data from the EOS instruments will be supplemented by measurements from operational sensors, most notably on satellites operated by NOAA. Some data centers will interoperate with the EOSDIS, allowing the DAACs and their users to search data inventories much as if the data were resident at one of the DAACs. Other data centers will not necessarily interoperate with the EOSDIS, but may provide data for the EOS science investigations.

2.3 ECS Requirements

The ECS will provide hardware, software and operations to:

a. plan, schedule, control, and command the space elements of the EOS mission;

b. receive, process, archive, and manage all data from EOS instruments and their associated data products;

c. receive, archive, and manage selected data from other NASA Earth Probe flight missions and other selected remotely sensed and *in situ* correlative data;

d. provide the Earth science community with access to all EOS data and other Earth science data held by the ECS and the data products resulting from research using these data;

e. promote effective utilization of data for research in support of the MTPE goals, by encouraging exchange of data and research results within the science community and across the multi-agency / multi-national data collection systems and archives; and

f. facilitate development, experimental usage, and ultimately community acceptance of new and/or improved algorithms for computing geophysical parameters from remotely sensed data and documenting the state of the Earth system.

These functions are classified as "mission critical", "mission essential", or "mission success". The ECS design must provide for all functions, but the implementation strategy will vary across these three classes of functions. Mission critical functions must be implemented on a robust and predictable schedule to support launches and ensure that data are not lost. At the other end of the spectrum, mission success functions will start out with proof-of-concept prototypes, and be implemented in the operational system as technical feasibility is demonstrated and within overall program schedule and budget constraints.

Item a describes mission critical functions for managing the EOS space assets. If these functions are not successfully performed, the remaining items are largely irrelevant and a major opportunity to document the state of the Earth system will be irrevocably lost. Fortunately, spacecraft and instrument control are traditional aerospace applications with established requirements, amenable to a schedule-driven development process. As in the previous section, we will not consider the flight operations aspects of the ECS.

Item b requires cost-effective processing with flexibility to incorporate new algorithms and data products as scientific understanding of the remote-sensing issues improves. Capturing the raw data and processing it to the level required to confirm data validity are mission critical functions for the same reasons as described in the previous paragraph. Downstream processing of higher level products is mission essential since recovery from processing errors or loss of data products can be accomplished by reprocessing from the lower level input data. Data processing requirements are discussed in Section 2.3.1, while data archiving and management requirements are described in Section 2.3.2. Item c is an extension of the EOS data management function to Earth science data from non-EOS sources (including new data products from operational instruments and data from the Earth Probe research satellite missions).

The ECS must facilitate access to the data holdings, giving users tools to search, locate, select, and acquire the data best suited to their science investigations, and in formats which make the data easy to use (item d in the above list). The ECS must also provide information and documentation on the data products, including data descriptions, algorithm documentation, and calibration/validation information. These data and information may be maintained by the ECS, the EOS investigators, or at other data centers. The challenge is to develop innovative methods to access this large and diverse data base. Search strategies and data access methods will evolve over time, in response to the changing information needs and utilization patterns of the research community. In many cases, data access and presentation methods will be developed by the EOS science community, to be incorporated into the ECS. Basic data search, subsetting and access services to the science community are considered mission essential, while advanced services specifically targeted at increasing the Earth science user's productivity are classified as mission success functions. Data access requirements are described in Section 2.3.3.

Finally, the last two items support the vision of a comprehensive information system incorporating DAAC-unique extensions to the ECS, interoperating with other agency data systems as part of the GCDIS, and enabling the science community to exchange data and algorithms directly, as service providers. These design drivers must be primary considerations at every stage of the ECS development. System-wide requirements for an operable, evolvable and extensible ECS are discussed in Section 2.3.4.

2.3.1 Standard Product Generation Requirements

ECS facilities at the DAACs will produce Standard Products (including quicklook, browse, and metadata) based on algorithms developed by EOS investigators at their SCFs.[2] Software tools are being developed under the ECS contract to allow SCFs to simulate the DAAC production environment. The DAACs will support EOS investigators and other science users in the development and testing of algorithms, migration of algorithms from development into operations, and quality control of data products to meet scientists' needs. The DAACs are responsible for verifying that the algorithms function properly in the operational environment; the algorithm developers are responsible for scientific validation of the data products.

The DAACs will generate Standard Products by routine processing and/or on demand, in response to user requests. The scientific algorithms will have access to commonly used data sets such as digital terrain maps, land/sea areas, climatology, and operational meteorological analysis products. Quality assessment profiles for each Standard Product will be generated in coordination with scientists at the associated SCFs. Table 2-2 shows the schedule for making routinely processed Standard Products available to users.[3] In parallel with routine operations, the DAACs will integrate and test new algorithms to produce additional Standard Products, and modifications to existing algorithms. Some algorithms may be temporarily run in a mode where routine generation is not guaranteed, producing prototype products for evaluation and testing. Based on priorities established by the science community, the DAACs will reprocess all or portions of data sets to provide improved and consistent long-term data sets.

ECS Standard Product capacity is based on estimates for data volumes and at-launch processing loads provided by the science algorithm development teams, and assuming that these requirements will expand by 20% each year after launch. Each DAAC will be able to accommodate its assigned routine processing, with an equal capacity to support algorithm testing and other non-routine activities, and will also be able to reprocess data at twice the rate of the routine processing.[4] Sustained CPU processing rate is assumed to be 25% of the peak processing rate.

The ECS will be designed to allow replacement of processors with minimal impact on the science algorithm software. The ECS design and implementation will have the flexibility to accommodate expansion by up to a factor of three in the processing capacity with no changes in

---

2 The term "algorithm" refers to software developed by an investigator to be used in the generation of science products. Algorithms include source code, processing control scripts, and documentation.

3 The schedule in Table 2-2 assumes that initial on-orbit check out of the instruments and algorithms have been completed. This initial check out is expected to require approximately 90 days after launch.

4 Assume that reprocessing of all data is initiated at the end of the 90-day post-launch check out period. With reprocessing performed at twice the rate of the initial (routine) processing, reprocessing will be completed at 4.5 months (= 3 months + 45 days) after launch. If reprocessing is initiated each time that previous reprocessing catches up to the routine processing stream, then sufficient capacity is available to reprocess the data 11 times within the planned system lifetime. This assumes that the reprocessing capacity is fully utilized, at all times.

design and up to a factor of ten without major changes to the design. In other words, the cost of processing growth should scale linearly with processing requirements, within these limits.

Table 2-2. Availability of Products from Routine DAAC Processing

Product Type	Definition	Availability*
Quick-look	Up to 1% of the instrument data, designated for special processing to support instrument operations, event monitoring and field campaigns	1 hour
Quick-look	Up to 10% of the instrument data, designated for special processing to support instrument operations, event monitoring and field campaigns	6 hours
Level 0	Raw instrument data at original resolution, time ordered, and with duplicate packets removed	24 hours
Level 1	Radiometrically corrected and calibrated instrument data, in physical units	24 hours
Level 2	Retrieved geophysical variables at the location and similar resolution as the Level 1 source data	24 hours
Level 3	Data or geophysical variables that have been spatially and/or temporally resampled (*i.e.*, gridded data)	24 hours
Level 4	Model output and/or variables derived from lower level data which are not directly measured by the instruments	1 week
Metadata	Information about data provided by the data supplier or algorithm which may be used to select data for a particular scientific application	with Level 1-4 product
Browse Data	Subsetted, subsampled, and/or summarized data provided for users to rapidly examine product primary data set characteristics	with Level 1-4 product

* Time to generate and make data available to users, after all required inputs (including the lower level input products and ancillary data) are available. For example, Level 1 Standards Products will nominally be available to users within 48 hours after instrument data collection, assuming all other required inputs are received on a similar schedule.

2.3.2 Data Archive and Distribution Requirements

The DAACs will archive all input data required to support reprocessing of the Standard Products, unless these data are readily available from other sources. In addition to the Standard Products, associated metadata, browse data and documentation (for example, algorithms, processing histories, calibration data sets, and quality assessments/status) will be maintained. The metadata will be updated at receipt of quality assessments from either the Standard Product processing or from the SCFs. All data will be verified against the associated metadata by checking for the presence of required fields and formats, error-free input, correctness of the data file size, etc. The results of this verification will be added to the metadata, and sent to the data provider.

Specific DAACs will maintain the instrument historical data and status information, and the mission historical data and platform status/ephemeris information. In addition, the DAACs will archive the Pathfinder data sets, Earth Probe data, non-EOS correlative data used for validation of Standard Products, selected Special Products, and research results (research articles, algorithms, data sets, and software) from the EOS investigators. Each DAAC will have the capability of scanning or digitizing hardcopy input for the purpose of archiving documents.

(paragraph deleted)

At each DAAC, the archive and distribution system will routinely retrieve data to support processing/reprocessing, quality assessment by scientists at the SCFs, and in response to user requests (including standing orders for specific types of data). Logically grouped sets of data will be retrievable as a single entity. The archive and distribution system will perform data transformations with user specified algorithms and subsetting, subsampling or averaging data within a granule, based on defined criteria including geographical location, spectral band, and time.

The data will be distributed in standard data formats.[5] Where appropriate, data will be compressed prior to storage or distribution. Data will be distributed electronically or on physical media, based on user requests. Archive data requested for electronic transfer will be available to the network within an average of 2 minutes after receiving the request if no reformatting is required, and within an average of 5 minutes where a format conversion is requested. Archive data requested on physical media will be prepared for distribution within 24 hours of receiving the request. In the case of a request requiring subsetting, subsampling or other transformations of the data, users will be provided estimates of delivery times, and status updates upon request.

The archive and distribution system is designed to accommodate the estimated at-launch daily data volumes, with allowance for 20% per year growth after launch. Each DAAC will be capable of providing data for electronic distribution at a rate equal to the daily product volume, with equal capacity for distribution via physical media. Each DAAC will be capable of supporting a 200% growth in the above requirements without architecture or design changes.

2.3.3 Data Access Requirements

The EOSDIS will contain diverse data sets of potential use to a very broad range of users. Potential data users include the EOS investigators, non-EOS-affiliated science users, and other users, including: policy makers and implementors, external system and tool developers, small data set users seeking illustrative data, users of EOS data for commercial and engineering applications, and educational users (including both educators and students). The combination of data diversity and user diversity requires that special attention be paid to developing effective data search and access strategies.

The ECS will provide users with information and tools to search, locate, select, and access data products. Users will also be able to access system, mission, instrument and algorithm

[5] NASA has selected the Hierarchical Data Format (HDF), developed by the National Center for Supercomputing Applications at the University of Illinois, as the standard format for distribution of EOSDIS data. However, other data distribution formats will also be supported.

documentation, publications, instrument and platform historical data, platform housekeeping and ancillary information, data acquisition and processing schedules, accounting information, and the status of standing data orders. The ECS will also maintain and provide access to a science processing library (software developed by the science community).

Users will access the ECS through the DAAC associated with their discipline, providing access to system-wide services spanning the whole of the EOSDIS including all DAACs, on-line SCFs, and external data centers. Data access services will be available on a 24 hour per day, 7 day per week basis, via direct connection, dial-up connection or a network link. Client-server, dumb terminal, and process-to-process access modes will be supported. Field support terminals will provide access through one or more of the DAACs to data at field experiment sites.

New users will complete on-line account applications and down-load client software for accessing EOSDIS services. (Limited services will also be available for unregistered users.) A user interface will outline available services with on-line help facilities to assist in using these services. The user interface will accommodate a wide range of user expertise and be adaptable to each user's profile. The interface will guide users to easily construct queries to identify desired data without detailed knowledge of the ECS architecture, data base management system, data base structure, query languages, or data formats. Users will be able to perform incremental searches by refining or adding criteria to identify the data of greatest interest.

The EOSDIS metadata is being designed to support complex multi-disciplinary and multi-instrument data searches by both novice and experienced users. An Earth science master directory will describe data sets available from the EOSDIS (including the SCFs), International Partner data centers, and external data centers. A guide of documentation and reference materials (including graphics and hypertext) for each EOSDIS data set will provide references to algorithms, data quality assessments, available literature, product/instrument specifications, summaries of related data sets, and user options available with respect to data transformations. Inventory-level metadata will describe individual granules, where a granule is the minimum logical unit of data stored in the archives. These metadata will be expandable to accommodate additional data attributes which are later determined to be useful for data search.

Users will search for data using combinations of: geographic coordinate reference, bounding polygon, point and radius, common geographic names, and data content (based on statistical summaries in the metadata). Searches for data coincident in time and/or space across data sets at distributed archives will be supported. Geographical and geophysical overlays will be used to display data coverage. Non-geographic metadata will be searchable by combinations of: exact word match, phrase match, character string, wild card constructs, character range, logical and Boolean operators, and min/max range. Document text search will also be supported.

In addition to selecting data based on information in the above data bases, users will be able to browse subsetted, subsampled and summarized data sets. Browse data sets will be generated by algorithms provided by EOS investigators. The ECS will be able to create *ad hoc* data sets in response to user requests for subsetting, subsampling or averaging within a granule, based on temporal, geographical coordinates, and/or spectral band criteria. For browse products which require additional processing, the user will receive an estimate of when the product will be ready for examination.

Once data of interest have been identified, on-line procedures will assist users in accessing existing data products, requesting additional processing of existing data, and/or submitting requests for new data acquisitions by EOS instruments. It is anticipated that scientists requiring new data are not necessarily experts on the EOS instruments. The ECS will provide on-line assistance for associating data products and/or geophysical parameters with the instrument(s) to acquire the input data, as well as recommended and allowable instrument parameter settings. Geographical reference aids, spacecraft location projections, and general feasibility checks will assist users in formulating data acquisition requests. Users will have access to long- and short-term EOS schedules for data acquisition, and to the existing list of requests which have been submitted, but not yet scheduled.

For all types of requests, the user's profile will facilitate the process and ensure that the request is within the access privileges and resources of the investigator. For each request, the user will be provided estimates of the cost and schedule for data acquisition and/or data processing. Status will be updated periodically as the request is confirmed (or rejected, with reasons), scheduled, and at various stages of completion.

Specific criteria have been developed for maximum response times for various on-line services, assuming a user community of 100,000 registered users, with 100 simultaneous sessions distributed across the EOSDIS DAACs. Response times for complex operations (*e.g.*, multiple DAAC inventory searches, retrieval and display of browse products) will be less than 60 seconds, with shorter response times for simpler operations. The ECS is being designed to accommodate 100% growth in the user service load without major design changes, in addition to an anticipated 20% per year growth in the rate of accumulation of inventory entries.

The ECS will provide users software toolkits to support their interactions with the information system, data, and instruments (see Table 2-3). Data manipulation and visualization tools will support ingest, subsetting and examination of EOSDIS data products. Communications software will allow users to access ECS services in a variety of modes (*e.g.*, dumb terminal, X-terminal, client-server, or process-to-process), supporting file transfers and providing access to ECS bulletin boards and multi-media mail. Data management and information access tools will interact with local data base management systems to generate and maintain metadata, allow easy transfer of data and metadata to and from the ECS, and provide the SCFs with the ability to act as data and service providers for other EOSDIS users. Algorithm development tools are designed to permit scientists to develop algorithms which can be easily and reliably ported to the operational DAAC environment. Finally, instrument scientists must be able to participate in the planning, scheduling and monitoring of their instruments. Instrument support tools will allow scientists to exchange plans, commands and status information with the ECS flight operations facility and to examine engineering and quick-look data from their instruments. The configuration of tools will be tailorable to the user's site-specific requirements, with a common user interface for accessing the tools.

2.3.4 System-wide Requirements

The ECS will be a distributed system with multiple DAACs, each serving a particular community within the multi-disciplinary Earth system sciences. The DAACs will be tightly coupled by requirements to support interdependent data production (*i.e.*, processing at a given DAAC may require inputs from other DAACs) within the overall product delivery schedule previously identified in Table 2-2. DAAC local autonomy is essential to encouraging the diversity of services and competition which is the strength of a truly distributed system. Thus, data processing and other services must be scheduled by the local DAAC staff, in accordance with resource availability and policies. This local autonomy must be balanced against the need to maintain product delivery schedules where processing involves inter-site dependencies. The ECS will support system-wide monitoring and coordinate DAAC-to-DAAC data transfers to support the overall Standard Product schedule, with local system management performed at each DAAC. A performance management service will perform long-term trend analysis to identify needs for upgrading ECS components and adjusting operations to respond to changing usage patterns. A system-wide fault management service will locate, isolate and identify operational problems to coordinate recovery or corrective actions. Configuration control, security, accounting, and directory services will also be needed at the system level as required, for instance, to allow users to access all of the ECS data holdings and services from any DAAC entry point. Providing such services at the system level does not imply that the associated data (user access privileges, for instance) need be maintained at the system level.

Table 2-3. ECS Toolkit Functions

Application	Users	Functions Provided
data manipulation	all	format conversion of EOS data; data subsetting; data compression (lossy and lossless); data transformation; data subsampling
data visualization	all	two-dimensional plots (x-y plots, scatter plots, profiles, histograms); three-dimensional plots; contour plots; three-dimensional surface plots; pan and zoom; cursor positioning by Earth, data and display coordinates
communications	all	GOSIP/ISO and TCP/IP protocols; process-to-process communications; electronic mail/message handling; virtual terminal support; file transfer and management; window managers; directory services; network security and access control
data management	SCFs	application programmer interfaces to build / maintain SCF databases consistent with EOSDIS data schema and metadata definitions
information access	SCFs	query builders for data search functions
algorithm development	algorithm developers	file I/O; error/status reporting; process control; memory management; bit manipulation; ancillary data access and manipulation; spacecraft ephemeris and attitude data access; time and date conversion; celestial body position and coordinate transformations; math/statistics library; constants and unit conversions; graphics support
instrument support	instrument scientists	event notification; planning and scheduling; instrument/platform status display; quick-look data display; performance analysis; history log access; data acquisition request evaluation

The ECS architecture is based on modular components that may be extended by the addition of other components, or evolved by the replacement of components. Software components are designed to facilitate re-use by external systems developers, with configuration-controlled applications programmer interfaces (APIs) permitting extensions to the core system functionality. Interoperating external data centers will enhance the diversity of the EOSDIS by providing access to other collection systems and data sets. The diversity and richness of the data holdings will expand as the system matures, offering a unique resource for understanding the processes which govern the Earth's ecosystem and physical climate, as well as for new applications of global and regional environmental data sets. Extensions to the core system functionality and data will arise from at least three external sources: DAAC-unique functionality and their non-ECS data holdings; services and data provided by the science community; and other agency data systems interoperating as components of the GCDIS.

As part of the EOSDIS Version 0 predecessor to the ECS, each of the DAACs are developing services tailored to the particular needs of their users. Anticipated DAAC-unique extensions include new metadata fields, browse products, data product descriptions, and data types required to serve discipline-specific needs. Local user interfaces may replace the core system interface to more closely match community-specific needs for data search and access. Extensions to the algorithm development toolkit may be required to more effectively support algorithms for discipline-specific data products. DAAC-unique value added services (*e.g.*, browse and visualization, on-line analysis, unique data combinations) will require APIs for accessing the DAAC production environment, data file servers, data base servers, and system management services. Different data storage devices may be utilized to service DAAC-specific needs.

EOS science investigators are developing a wide range of Special Products at their SCFs. Some of these Special Products will be stored and distributed by the DAACs, but the SCFs must also be able to advertise their own data and services directly to the EOSDIS user community, through the same mechanisms that provide access to the DAAC data and services. Science community extensions to the EOSDIS capabilities are particularly anticipated in advanced data search methods that dynamically browse data and metadata. These will be developed at SCFs, supported by the ECS toolkits and by export of all or major subsets of the EOSDIS metadata to user computing facilities. When these dynamical search methods are interactively inserted at the DAACs, up to 10% of the ECS resources will be available for science community utilization, without interfering with routine data production, distribution and access functions.

The GCDIS is a multi-agency confederation of data systems interoperating to allow users from any participating agency to access data archived by any other participating agency. All the data acquired by Mission to Planet Earth will be available to the GCDIS and to its widespread community of global change researchers. Although the specific ECS requirements in support of GCDIS are still being defined, the ECS must be developed in a manner that enables cost-effective realization of the GCDIS objectives. The ECS architecture must not preclude separate installation of information management, data production, or data archive and distribution software components into other GCDIS data centers. (Re-use of information management functions, in particular, can enhance interoperability with these data centers.) Such re-use requires controlled and documented APIs between components serving different functions, allowing these components to be grouped and extended according to the needs of each data center. The ECS will be designed to accommodate GCDIS interoperability and component re-use, but implementation of other agency data systems will be implemented under their separate funding and control.

Finally, the ECS design must be able to incorporate new developments driven by the broader information technology community. The last few years has seen greatly expanded utilization of flexible directory search techniques, as in the Wide Area Information Server - WAIS (MacRae and Jones, 1992) and "gopher" (McCahill, 1992), and applications of hypertext to linking logically connected information, as in the World Wide Web - WWW (Berners-Lee *et al.*, 1992) and NCSA Mosaic (Andressen, 1993).[6] Likewise, the gigabit network test beds (Catlett, 1993) are likely to produce technologies directly applicable to the EOSDIS challenge of allowing users to interact with vast amounts of data and information. The ability to incorporate new technologies, as well as responding to changing user needs is critical if the information system is to remain viable over the planned multi-decade lifetime.

2.4 Development Process

Implementation of the ECS will be accomplished through an evolutionary development process which will allow graceful transition from the existing Version 0 system to the full-scale EOSDIS configuration. The goal of the evolutionary process is to develop the system that best matches user needs, even as these needs change. Key elements that support evolutionary development include: user involvement, prototypes, competition among potential solutions arising both within the project and from outside developments, ongoing technology assessments, encouraging innovation to provide new approaches and/or solutions, and multiple track development.

User involvement is a common thread through the evolutionary development process. User feedback on requirements, prototypes, and design involves formal reviews with participation by the EOSDIS Science Advisory Panel, ongoing analysis of user-sensitive issues by EOSDIS Focus Teams, ECS Tirekickers (funded researchers available for brainstorming and reviewing development concepts very early in the implementation process), tracking the user community through a User/Data Model, soliciting user comments through surveys and interviews, and collaborative prototyping with the ECS user community. A Recommended Requirements Data Base provides the mechanism for tracking user feedback from the initial requirements refinement through implementation. User satisfaction metrics provide a formal and quantitative method for measuring user reaction to fielded ECS prototypes and releases, which provide the basis for design of future releases.

Figure 2-3 shows the ECS development schedule, and illustrates the multiple track development process. The ECS will be developed in four releases (see following paragraph for individual release descriptions). Each release includes a formal development track, supplemented one or more incremental tracks. The formal development track produces high reliability products through a rigorous process of concept definition, system analysis, design, implementation, integration, and test. This is the preferred track for mission critical functions with design requirements which are well understood in advance, and where early and detailed interface specifications are required to permit parallel development of external systems such as the EOS

6 See Schwartz *et al.* (1992) for a summary and comparison of Internet information search techniques.

instruments. The incremental tracks consist of multiple deliveries (evaluation packages), typically separated by six months. This allows for iteration of design with extensive user inputs to design requirements. Documentation for the incremental development is less extensive than for the formal track, typically consisting of white papers, development notebooks, briefing charts, and demonstrations. Linkages between the incremental and formal tracks include rigorous definition of the interfaces between the two developments, and integration of the incremental deliveries into the formal release prior to Test Readiness Reviews.

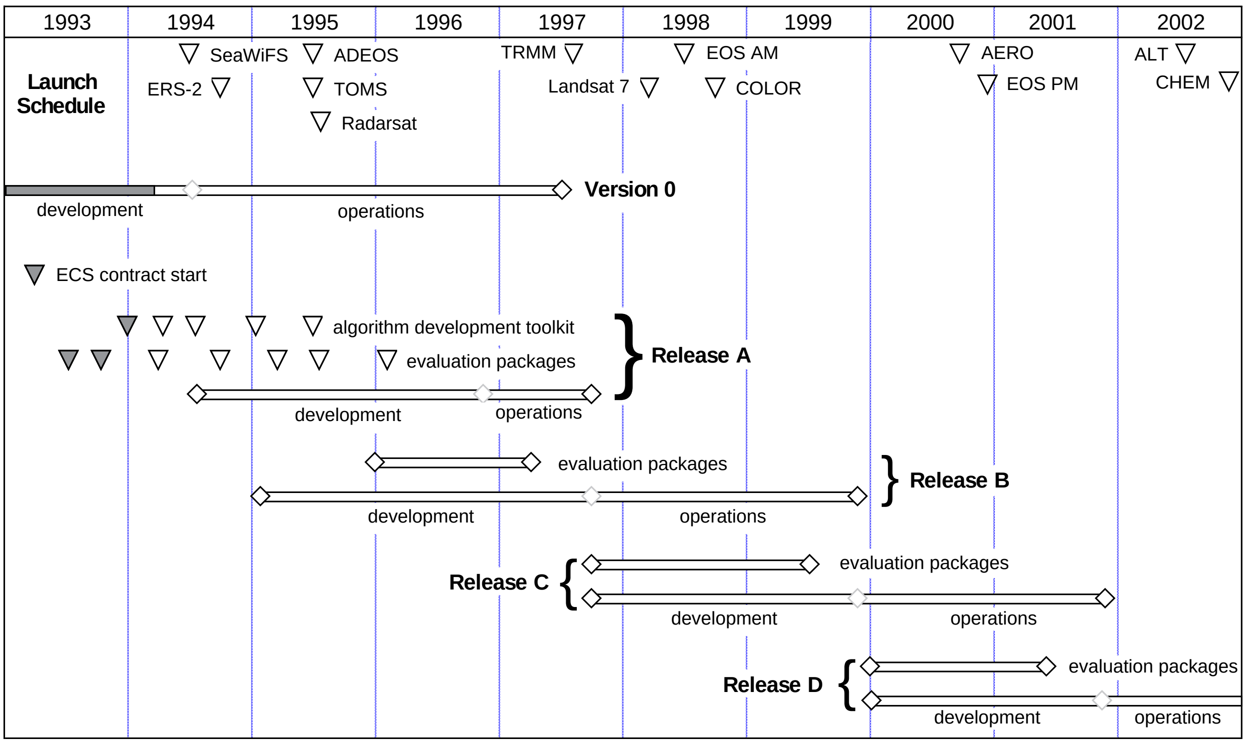

Figure 2-3. ECS Development Schedule

Release A will support data processing, archiving, and distribution for TRMM, Landsat 7, and COLOR, in addition to providing for flight operations interface testing and end-to-end data flow testing for EOS AM. In addition to evaluation packages and a formal track, Release A includes the algorithm development toolkit to support transition of algorithm software developed by the science community into the ECS DAACs. Release A will also include an Interim Release (scheduled for the end of 1995, but not shown in Figure 2-3) to support early interface testing for TRMM, Landsat 7 and EOS AM. Release B will provide flight operations and data functions for the EOS AM platform and instruments, which will be expanded to include AERO and EOS PM in Release C, and ALT and CHEM in Release D. Release A will interoperate with Version 0 to provide bi-directional access to each system's data holdings, with the Version 0 data migrated to the ECS in Release B. Successive releases will provide expanded and increasingly sophisticated data search and access, based on feedback from the science community and capabilities demonstrated by external prototypes and Evaluation Packages.

In addition to developing the ECS in an evolutionary manner, the delivered system must be evolvable. System evolvability is the ability of the delivered system to economically adapt to future changes whether those changes are in user paradigm or technology. In order to develop an evolvable system, the ECS will be analyzed relative to predicted future changes. Evolvability requirements and evolvability tests are developed as predictions of future changes. Evolvability requirements are those relatively near-term changes which ECS must enable by design. "Enable" means ECS will be packaged and documented to allow others to execute the change, perhaps at a later time. Evolvability tests are those relatively long-term changes which ECS must not preclude. Evolvability tests provide the basis for analyses which determine the flexibility of the ECS architecture to change. By defining an evolvable architecture along with implementing evolvability requirements and tests, the ECS is prepared for system evolvability.

2.5 Summary and Further Reading

This white paper summarizes the ECS requirements from a science-user perspective. It was developed in response to feedback from the science community that a concise summary of planned ECS functionality is needed to facilitate user feedback. The formal contractual specifications describe the system functional and performance requirements by decomposing these requirements into a long list of testable statements. This is necessary to determine that the delivered system complies with the terms of the contract, but it does not lead to a highly readable document. These contractual specifications are contained in:

> *Functional and Performance Requirements Specification for the Earth Observing System Data and Information System (EOSDIS) Core System*, NASA Goddard Space Flight Center, February 1993.

These specifications are currently being revised based on feedback from the science and information systems communities. The proposed changes are enumerated in:

> *EOSDIS Core System (ECS) Requirements Specification*, ECS Document 194-216-SE1-001, February 1994,

which allocates requirements to the elements identified in the previous document, but is not considered to specify the final ECS architecture and design.

The ECS project is also producing a series of white papers, as informal technical documents addressing specific topics. In particular, further information on the evolutionary development process is offered by:

> *Multi-Track Development for the ECS Project*, ECS White Paper FB9404V2, March 1994.

The requirements presented in Section 2.3 apply over the lifetime of the program. Specific capabilities to be provided by each of the ECS Releases are proposed in:

> *Release Plan Content Description*, ECS White Paper FB9403V1, February 1994.

Key concepts in the design of the information system architecture are described in:

> *EOSDIS Core System Science Information Architecture*, ECS White Paper FB9401V2, March 1994.

These documents may be obtained from the ECS Data Management Office (see Section 1.3, above), which can also provide a complete list of currently available ECS white papers.

As is appropriate for an evolutionary system, the planned ECS functionality is being periodically revised in response to evolving science needs (as well as improved understanding of these needs), emerging information technology opportunities, and operational experience with Version 0 and other science data systems. In this fluid (but cost constrained) environment, it is necessary to establish priorities in order to develop an effective and affordable information system. Understanding of cost implications is required in order to make intelligent choices, but specific cost sensitivity is design dependent. Accordingly, a white paper will provide an analysis of cost sensitivities, as a science-oriented summary of the System Design Specification (scheduled for June 1994).

Communications between the diverse ECS user communities, system developers, and policy makers is (and will remain) a challenge which requires the ongoing attention of all the stakeholders. It is hoped that this white paper (and similar papers referenced above) will contribute to meeting this challenge by providing the user community a window into the ECS development plans and, thus, the opportunity to influence these plans.

2.6 References

Andressen, M., "NCSA Mosaic Technical Summary", Technical Report 2.1, Software Development Group, National Center for Supercomputing Applications, Champaign, Illinois, May 1993.

Asrar, G. and Dokken, D.J. (ed.), *EOS Reference Handbook*, NP-202, 1993.

Berners-Lee, T., Cailliau, R., Groff, J.-F., and Pollerman, B., "World-Wide Web: The Information Universe", *Electronic Networking: Research, Applications and Policy*, vol. 2, pp. 52-58. April 1992.

Butler, D.M. *et al.*, *Earth Observing System: Science and Mission Requirements Working Group Report*, Vol. I & Appendix, NASA TM-86129, 1984.

Catlett, C.E., "Internet Evolution and Future Directions", in *Internet System Handbook*, ed. D.C. Lynch and M.T. Rose, Addison-Wesley Publishing Company, 1993.

Chase, R.R.P. *et al.*, *Earth Observing System Data and Information System: Report of the EOS Data Panel*, NASA TM-87777, Vol. IIa, 1986.

MacRae, T., and Jones, S., "WAIS: A New Development in Information Services", *MIT Information Science Newsletter*, June 1992.

McCahill, M., "The Internet Gopher: A Distributed Server Information System", *ConneXions - The Interoperability Report, Interop*, Inc., vol. 6, no. 7, pp. 10-14, July 1992.

Schwaller, M., and Andrews, M., *Science Data Plan for the EOS Data and Information System, Covering EOSDIS Version 0 and Beyond*, Document Version 2.0, NASA Goddard Space Flight Center, June 1993.

Schwartz, M.F., Emtage, A., Kahle, B., and Neuman, C.B., "A Comparison of Internet Resource Discovery Approaches", *Computing Systems (The Journal of the USENIX Association)*, vol. 5, pp. 461-493, Fall 1992.

This page intentionally left blank.

Abbreviations and Acronyms

ADEOS	Advanced Earth Observing Satellite (Japan)
API	applications programmer interface
AVHRR	Advanced Very High Resolution Radiometer
CERES	Clouds and Earth's Radiant Energy System
DAAC	Distributed Active Archive Center
DMSP	Defense Meteorological Satellite Program
ECS	EOSDIS Core System
EDC	EROS Data Center
EOS	Earth Observing System
EOSDIS	EOS Data and Information System
EROS	Earth Resources Observation System
ERS	European Remote-sensing Satellite
ESA	European Space Agency
GCDIS	Global Change Data and Information System
GOES	Geostationary Operational Environmental Satellite
GOSIP	Government Open System Interconnection Profile
GSFC	Goddard Space Flight Center
IR	infrared
ISO	International Standards Organization
JERS	Japanese Earth Remote-sensing Satellite
JPL	Jet Propulsion Laboratory
Landsat	Land Remote Sensing Satellite
LaRC	Langley Research Center
LIS	Lightning Imaging Sensor
MSFC	Marshall Space Flight Center
MTPE	Mission to Planet Earth
NASA	National Aeronautics and Space Administration
NASDA	National Space Development Agency of Japan

NCSA	National Center for Supercomputing Applications
NOAA	National Oceanic and Atmospheric Administration
NSIDC	National Snow and Ice Data Center
ORNL	Oak Ridge National Laboratory
Radarsat	Synthetic Aperture Radar Satellite (Canada)
SAR	Synthetic Aperture Radar
SCF	Scientific Computing Facility
SeaWiFS	Sea-viewing Wide Field-of-view Sensor
SSM/I	Special Sensor Microwave/Imager
TCP/IP	transport control protocol / Internet protocol
TDRS	Tracking and Data Relay Satellites
TIROS	Television and Infrared Observation Satellite
TOMS	Total Ozone Mapping Spectrometer
TOPEX	Ocean Topography Experiment (U.S.-France)
TOVS	TIROS Operational Vertical Sounder
TRMM	Tropical Rainfall Measuring Mission (U.S.-Japan)
UAF	University of Alaska - Fairbanks
UARS	Upper Atmosphere Research Satellite
vis	visible
WAIS	Wide Area Information Server
WWW	World Wide Web